Report

Complex Segregation Analysis Provides Compelling Evidence for a Major Gene Underlying Obsessive-Compulsive Disorder and for Heterogeneity by Sex

G. Nestadt,[1] T. Lan,[3,6] J. Samuels,[1] M. Riddle,[1,2] O. J. Bienvenu III,[1] K. Y. Liang,[4] R. Hoehn-Saric,[1] B. Cullen,[1] M. Grados,[1] T. H. Beaty,[3] and Y. Y. Shugart[2,5]

Departments of [1]Psychiatry and Behavioral Sciences and [2]Pediatrics, School of Medicine, and Departments of [3]Epidemiology and [4]Biostatistics, School of Public Health, and [5]Center for Inherited Disease Research, The Johns Hopkins University, Baltimore; and [6]Department of Psychiatry, Yu-Li Veterans Hospital, Taiwan

Evidence from twin and family studies supports a genetic etiology for obsessive-compulsive disorder (OCD). The purpose of this study was to test whether a major gene is implicated in a proportion of families with OCD. Complex segregation analyses of 153 families (80 case and 73 control), ascertained in the Johns Hopkins OCD Family Study, provided strong evidence for a major gene. A Mendelian-dominant model, with significant sex effects and with residual familial effects, best explained the observed data. Stratification of the sample by the sex of probands provided further evidence of heterogeneity with respect to familial aggregation. Segregation analyses of 86 families with a female proband and of the 67 families with a male proband suggested that a Mendelian-dominant model with familial residual effects was the most parsimonious model explaining the inheritance of OCD in both subgroups.

Obsessive-compulsive disorder (OCD [MIM 164230]) is a psychiatric disorder affecting 1%–3% of the population (Samuels and Nestadt 1997). There is substantial evidence that it is a genetic condition. Twin studies have shown substantial concordance among MZ (70%–80%) compared with DZ twin pairs (22%–47%) (Inouye 1965; Carey and Gottesman 1981). Recent family studies found a significant increase in the risk of the disorder in relatives of affected probands compared with relatives of controls (Pauls et al. 1995; Nestadt et al. 2000).

The mode of inheritance of OCD has been investigated by means of segregation analysis in three studies. Evidence of a gene of major effect was found in two studies (Nicolini et al. 1991; Cavallini et al. 1999); however, a more precise pattern of inheritance could not be established. Alsobrook et al. (1999) studied inheritance patterns in a sample of families with OCD, using symptom-based factor scores. In a subset of families with higher

scores on a specific symptoms factor—symmetry and ordering—the polygenic model was rejected, which suggests the existence of a major locus.

To further explore the mode of inheritance of OCD in the U.S. population, we ascertained 80 case families (423 subjects, including probands and their first-degree relatives), and 73 control families (373 subjects, including probands and their first-degree relatives). Sample ascertainment and diagnostic procedures have been described elsewhere (Nestadt et al. 2000). In brief, 80 adult probands with OCD who meet strict DSM-IV (American Psychiatric Association 1994) criteria for OCD were recruited from five specialty OCD treatment centers in the Baltimore/Washington, DC, area. Control probands were ascertained through a random-digit–dialing procedure and were matched to case probands on sex, race, and age. Probands were excluded from the study if they were diagnosed with schizophrenia, mental retardation, dementia, or Tourette disorder or if OCD occurred exclusively during a major depressive episode.

After informed consent was obtained, a semistructured interview was conducted by a clinician blinded to the subjects' proband status and to whether they were from a case or control family. The primary diagnostic instrument was the Schedule for Affective Disorders and

Schizophrenia-Life Time Anxiety (SADS-LA) (Mannuzza et al. 1986). Children aged 8–15 years were evaluated using the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS [Kaufman et al. 1997]). In addition to information from the direct interview, family-informant interviews were conducted for diagnostic purposes. All this information, as well as all available hospital records, was made available to two psychiatrists, each of whom independently reviewed the materials and together arrived at a consensus diagnosis under strict DSM-IV criteria.

There were more males in the case families than in the control families (51% vs. 44%; $P = .046$); the members of case families were slightly older at the time of interview (mean 45.9 years vs. 43.3 years; $P = .038$); and, if affected, they had an earlier age at onset (mean 11.8 years vs. 17.8 years; $P = .06$). As would be expected, the prevalence of definite OCD was significantly greater among case families than among control families (29.1% vs. 2.1%; $P < .00001$).

In the 80 case families, males and females did not differ significantly with respect to the proportion diagnosed with OCD (29.5% vs. 26.4%; $P = .48$), their mean age at interview (~46 years), or their mean age at onset of OCD (~12 years). Subjects ascertained through either a male or a female proband had similar mean age at interview (~45 years), age at OCD onset (~12 years), and proportion diagnosed with OCD (~16%).

Complex segregation analysis was performed using the program REGD, as implemented in the S.A.G.E. package (version 3.1) (1997). This program uses regressive logistic models (Bonney 1986) to test for the presence of a major susceptibility locus, residual correlations in risk among related individuals, and the effect of measured risk factors. In this approach, a regression relationship is formulated such that the phenotype of a person is dependent on an unobserved "type" and other measured covariates. The concept of "type" (Go et al. 1978) was used to describe the discrete factors that affect an individual's phenotype when testing genetic hypotheses. In the current analysis, these types were modeled so that they were dependent on the types of preceding relatives; therefore, likelihood-based tests of significance could be constructed to evaluate the role of underlying genetic factors. The regressive model (class A) was formulated as follows:

$$\text{Log} \frac{P(\text{affected})}{1 - P(\text{affected})} = \beta_i + \delta_{sp}(\text{spouse affected})$$
$$+ \delta_{po}(\text{parents affected})$$
$$+ \xi(\text{covariates}) ,$$

where P(affected) is the risk of having the OCD phe-

notype; β_i is the baseline risk for three types; δ_{sp} and δ_{po} represent, respectively, the residual effects of having an affected spouse and an affected parent; and ξ indicates the risk explained by the other observed risk factors (covariates). Single-ascertainment correction was applied (Cannings and Thompson 1977).

The following series of competing models was tested: a sporadic model; Mendelian models including dominant, recessive, and codominant; and an environmental model. The likelihood of a general unrestricted model was calculated and compared with the specified reduced models, with one or more relevant parameters restricted. To test hypotheses regarding specific models of inheritance, $-2\ln L$ of the general model was subtracted from $-2\ln L$ of the reduced model of interest. This difference is distributed asymptotically as a χ^2 with $n–k$ df, with n and k being, respectively, the number of parameters estimated in the general model and in the reduced model.

The most parsimonious model, with a log-likelihood not significantly different from the most general model, was identified using Akaike's Information Criteria (AIC) (Akaike 1974), which is defined as AIC = -2(ln likelihood) + 2(no. of parameters).

This AIC serves as a weighted measure of the fit of any given model. It could also serve as a guide for the best-fitting model when two models are not nested in their parameter space.

Table 1 presents the estimated parameters for six models of inheritance fit to the entire sample of 153 families (composed of 80 case families and 73 control families). Comparisons of all the restricted models with the most general model (model 6) showed that neither the Mendelian dominant model (model 2) ($\chi^2_5 = 6.61$, $P = .25$) nor the Mendelian codominant model (model 4) ($\chi^2_3 = 6.76$, $P = .08$) could be rejected on the basis of the likelihood ratio test (LRT). The Mendelian recessive model (model 3) was rejected ($\chi^2_5 = 13.75$, $P = .017$). All models omitting residual familial effects (sporadic, dominant, recessive, and codominant models) were rejected ($P < .001$ for each model). This suggested that, in addition to the major locus, unexplained familial effects had an important role in the expression of the OCD phenotype. The sporadic model (model 1) and the environmental model (model 5) were strongly rejected ($P < .0001$). Comparing the AIC scores for models 2 and 4 indicated that the Mendelian dominant model (model 2) was the most parsimonious explanation for the inheritance of OCD in the total sample. However, the LRT, comparing dominant and codominant models, was not significant.

Estimates from model 1 indicated that 0.2% of subjects were homozygous for the high-risk genotype (AA), 8.6% were heterozygous carriers (AB), and 91.2% had the low-risk genotype (BB). The sex-specific genetic penetrance was 86.7% for females and 78.5% for males,

Table 1

Segregation Analysis (REGD Class A) of Definite OCD Phenotype in 776 Members of 153 Pedigrees, after Ascertainment Adjustment on the Basis of Proband Status

Model	$P(A)$	β(AA) Female	β(AB) Female	β(BB) Female	β(AA) Male	β(AB) Male	β(BB) Male	τ(AA)	τ(AB)	τ(BB)	δ_{vp}	δ_{po}	$-2\ln(L)$	AIC	N^c	χ^2 (df); P
1. Sporadic	(1.0)	−1.839±.182	β(AA) female	β(AA) female	−1.942±.200	β(AA) male	β(AA) male	...	...	...	.978±.40	−.312±.106	665.96	673.96	4	40.74 (8); <.001
2. Dominant	.045±.014	1.878±1.001	β(AA) female	−4.099±.704	1.294±.819	β(AA) male	−4.289±.710	[1.0]^b	[.5]^b	[.0]^b	1.646±.91	−1.435±.352	631.82	645.82	7	6.61(5); .25
3. Recessive	.322±.04	2.373±1.191	β(BB) female	−4.844±1.141	1.733±1.130	β(BB) female	−5.087±1.163	[1.0]^b	[.5]^b	[.0]^b	2.284±1.18	−1.716±.566	638.96	652.96	7	13.75 (5); .017
4. Codominant	.045±.014	(10)^a	1.842±1.041	−4.083±.700	(10)^a	1.269±.849	−4.270±.705	[1.0]^b	[.5]^b	[.0]^b	1.607±.91	−1.428±.350	631.97	649.97	9	6.76 (3); .079
5. Environmental	.138±.09	2.504±9.801	.109±1.440	−5.115±4.239	.169±13.424	.169±1.850	(−10)^a	$=q_A$	$=q_A$	$=q_A$	1.948±1.14	−.597±.508	665.02	683.02	9	39.81 (3); <.001
6. General	.00001±.00	−.063±1.018	−1.094±.405	(−10)^a	(10)^a	−2.052±.612	(−10)^a	(1.0)	(1.0)	.152±.032	.393±.65	−.360±.192	625.21	649.21	12	...

[a] Parameter value became fixed at a preset bound during likelihood maximization.

[b] Value was fixed by the specified Mendelian model.

[c] Number of parameters.

in the combined group of high-risk individuals (8.8%). This finding is consistent with that reported by Cavallini et al. (1999). Subjects who had a spouse with OCD had a 5.2-fold increased risk of OCD, compared with those who did not have a spouse with OCD, after the correction for the major-gene effect.

The combined penetrance for female subjects with affected spouses and high-risk genotypes (AA or AB) was 97.1%. The penetrance for female subjects with high-risk genotypes (AA or AB) and an affected parent was 60.9%. The penetrances for male subjects under the same conditions were 94.9% and 46.5%, respectively.

To test for heterogeneity between families ascertained through male and female probands, we compared the 86 families selected through a female proband and the 67 selected through a male proband. A heterogeneity χ^2 test (Williams and Anderson 1984; Khoury et al. 1993) compared the sum of $-2\ln L$ of a particular model, computed on the subsets, with the $-2\ln L$ computed on the total 153 families. This statistic is computed as follows: $\chi^2 = -2 \, [\Sigma \ln L(\text{best-fitted model/subgroup } i) - \ln L(\text{best-fitted model/all family data})]$, where Σ is the sum over all i subgroups. This statistic approximates a χ^2 with df equal to $[(I*K) - K]$ for $i = 1, 2, \ldots, I$ subgroups and a model with K parameters.

The best-fitting model (model 2 in table 1) was evaluated separately in these two subgroups. The $-2\ln L$ was 320.95 for the 86 families with a female proband and 261.69 for the 67 families with a male proband. The difference between the $2\ln L$ of model 2 and the sum of the $-2\ln L$ for the two groups yielded an LRT = 49.18 (heterogeneity χ^2 with 7 df [$P < .0001$]). This indicated strong evidence for etiologic heterogeneity between families ascertained through female probands and male probands.

The evidence for heterogeneity between families ascertained through female and male probands motivated us to perform separate segregation analyses of the 86 families with a female proband and of the 67 families with a male proband. In the analysis of the families with a female proband, the results of all models were very similar to those in the total sample. Both the sporadic and the environmental models were strongly rejected. Among the three Mendelian models, the recessive model was rejected ($P = .015$), but neither the dominant model nor the codominant model could be rejected. On the basis of the AIC (334.95 vs. 338.93), the dominant model provided a more parsimonious fit than the codominant model. Focusing on the best-fitting dominant model, the baseline risk of the high-risk genotypes (AA and AB) for females increased from 6.54 to 24.5, whereas the risk for males remained ~3.7.

In the segregation analysis of the 67 families with a male proband, the sporadic and the environmental models were strongly rejected. However, we could not dif-

ferentiate between the three Mendelian models because each fit as well as the general model. It was also difficult to choose between these models, using the AIC (the dominant model had an AIC of 275.7, the recessive an AIC of 276.8, and the codominant an AIC of 278.1). It was clear that this subgroup of families was compatible with a Mendelian major-locus model, but the details of the model were less evident.

The results from the total sample and the stratified subgroups were consistent with Mendelian inheritance of a dominant allele leading to a high risk of having OCD. However, given the sample size of this study, we could not rule out a codominant model.

In other studies (Nicolini et al. 1991; Cavallini et al. 1999), a major-gene effect was supported, but all dominant, additive, or recessive models could not be rejected. Our results, based on a relatively large, well-characterized sample, with only OCD as the affected phenotype, provide stronger evidence for an autosomal dominant or a codominant pattern of inheritance. Another strength of this study is that the sample included both case and control families; thus, the estimated disease frequencies and penetrances from this study are likely to be less biased.

The evidence of genetic heterogeneity on the basis of sex of the proband may suggest different genetic/environmental exposures in at least a proportion of families with OCD. Interestingly, this finding corresponds with reports in the literature that there is a slight preponderance of females with OCD (Karno et al. 1991). Another sex difference is earlier age at onset in males (Lensi et al. 1996; Nestadt et al. 1998). Additionally, obsessive-compulsive personality disorder, often considered related to OCD, tends to be more common among males (Nestadt et al. 1991). The sample size, in this study, limits our ability to address this issue further.

Only strict DSM-IV OCD was included in this study. Other disorders, if genetically related, may constitute alternate phenotypic expressions of the same underlying gene(s). These disorders include subsyndromal OCD symptoms (Nestadt et al. 2000), Tourette and other tic disorders (Pauls et al. 1986), eating disorders (Cavallini et al. 2000), obsessive-compulsive personality disorder (Samuels et al., 2000), body dysmorphic disorder (Bienvenu et al., 2000), and anxiety disorders (Nestadt et al., in press). Including these conditions as affected phenotypes could further strengthen the findings and may provide an explanation for the sex differences reported. For instance, Cavallini et al. (2000) reported that when they included OCD as the affected phenotype they found evidence of a Mendelian mode of inheritance in a family study of probands with eating disorders.

Results of the present study also showed that, in addition to the major-locus effect, it appears that there is a residual effect of an affected spouse. This effect may

indicate some degree of assortative mating, which is not elsewhere described in individuals with OCD.

In conclusion, the results strongly support a Mendelian dominant or codominant susceptibility gene for OCD, acting in a proportion of families. Nevertheless, Mendelian factors alone are not sufficient to fully explain the familial aggregation of this phenotype, and residual familial effects are necessary to adequately fit the data. This suggests that polygenic factors may also contribute to the etiology of OCD. Parameters derived from this study may facilitate future linkage studies.

Acknowledgments

Electronic-Database Information

The accession number and URL for data in this article are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for OCD [MIM 164230])

References

Akaike HA (1974) New look at the statistical model identification IEEE. Transmission Automation Control 19:716–723

Alsobrook JP II, Leckman JF, Goodman WK, Rasmussen SA, Pauls DL (1999) Segregation analysis of obsessive-compulsive disorder using symptom-based factor scores. Am J Med Genet 88:669–675

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th ed. American Psychiatric Association, Washington, DC

Bienvenu OJ, Samuels JF, Riddle MA, Hoehn-Saric R, Liang K-Y, Cullen B, Grados MA, Nestadt G (2000) The relationship of obsessive-compulsive disorder to possible spectrum disorders: results of a family study. Biol Psychiatry 48:287–293

Bonney GE (1986) Regressive logistic models for family disease and other binary traits. Biometrics 42:611–625

Cannings G, Thompson EA (1977) Ascertainment in the sequential sampling of the pedigrees. Clin Genet 12:208–212

Carey G, Gottesman II (1981) Twin and family studies of anxiety, phobic, and obsessive disorders. In: Klein DF, Rabkin JG (eds) Anxiety: new research and changing concepts. Raven Press, New York, pp 117–136

Cavallini MC, Bertelli S, Chiapparino D, Riboldi S, Bellodi L (2000) Complex segregation analysis of obsessive-compulsive disorder in 141 families of eating disorder probands, with and without obsessive-compulsive disorder. Am J Med Genet 96:384–391

Cavallini MC, Pasquale L, Bellodi L, Smeraldi E (1999) Complex segregation analysis of obsessive compulsive and spectrum disorders. Am J Med Genet 88:38–43

Go RC, Elston RC, Kaplan EB (1978) Efficiency and robustness of pedigree segregation analysis. Am J Hum Genet 30:28–37

Inouye E (1965) Similar and dissimilar manifestations of obsessive-compulsive neurosis in monozygotic twins. Am J Psych 121:1171–1175

Karno M, Golding JM (1991) Obsessive compulsive disorder. In: Robins LN, Regier DA (eds) Psychiatric disorders in America. Free Press, New York, pp 204–219

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N (1997) Schedule for affective disorders and schizophrenia for school-age children: present and lifetime version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 36:980–988

Khoury MJ, James LM (1993) Population and familial relative risks of disease associated with environmental factors in the presence of gene-environment interaction. Am J Epidemiol 137:1241–1250

Lensi P, Cassano GB, Correddu G, Ravagli S, Kunovac JL, Akiskal HS (1996) Obsessive-compulsive disorder: familial-developmental history, symptomatology, comorbidity and course with special reference to gender-related differences. Br J Psychiatry 169:101–107

Mannuzza S, Fyer AJ, Klein DF, Endicott J (1986) Schedule for affective disorders and schizophrenia-lifetime version modified for the study of anxiety disorders (SADS-LA): rationale and conceptual development. J Psychiatr Res 20:317–325

Nestadt G, Bienvenu OJ, Cai G, Samuels J, Eaton WW (1998) Incidence of obsessive-compulsive disorder in adults. J Nerv Ment Dis 186:401–406

Nestadt G, Romanoski AJ, Brown CH, Chahal R, Merchant A, Folstein MF, Gruenberg EM, McHugh PR (1991) DSM-III compulsive personality disorder: an epidemiological survey. Psychol Med 21:461–471

Nestadt G, Samuels JF, Riddle M, Bienvenu OJ, Liang KY, LaBuda M, Hoehn-Saric R, Walkup J, Grados M (2000) A family study of obsessive-compulsive disorder. Arch Gen Psychiatry 57:358–363

Nestadt G, Samuels J, Riddle MA, Liang K-Y, Bienvenu OJ, Hoehn-Saric R, Grados M, Cullen B. The relationship between obsessive-compulsive disorder and anxiety and affective disorders: results from the Johns Hopkins OCD family study. Psychol Med (in press)

Nicolini H, Hanna G, Baxter L Jr, Schwartz J, Weissbacker K, Spence MA (1991) Segregation analysis of obsessive compulsive and associated disorders: preliminary results. Ursus Medicus 1:25–28

Pauls DL, Alsobrook JP, Goodman W, Rasmussen S, Leckman JF (1995) A family study of obsessive-compulsive disorder. Am J Psychiatry 152:76–84

Pauls DL, Leckman JF (1986) The inheritance of Gilles de la Tourette syndrome and associated behaviors. N Engl J Med 315:993–997

S.A.G.E. (Statistical Analysis for Genetic Epidemiology), version 3.1 (1997) Computer program package, available from the Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland

Samuels J, Nestadt G (1997) Epidemiology and genetics of obsessive-compulsive disorder. Int Rev Psychiatry 9:61–71

Samuels J, Nestadt G, Bienvenu OJ, Costa PT Jr, Riddle MA, Liang K-Y, Hoehn-Saric R, Grados MA, Cullen B (2000) Personality disorders and normal personality dimensions in obsessive-compulsive disorder: results from the Johns Hopkins OCD Family Study. Br J Psychiatry 177:457–462

Williams WR, Anderson DE (1984) Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. Genet Epidemiol 1:7–20